

Linear estimates for solutions of quadratic equations in free groups

Olga Kharlampovich and Alina Vdovina

July 15, 2011

Abstract

We prove that in a free group the length of the value of each variable in a minimal solution of a standard quadratic equation is bounded by $2s$ for orientable equation and by $12s^4$ for non-orientable equation, where s is the sum of the lengths of the coefficients.

1 Introduction

The study of quadratic equations over free groups started with the work of Malcev [13] and has been deepened extensively ever since. One of the reasons research in this topic has been so fruitful is the deep connection between quadratic equations and the topology of surfaces.

In [6] the problem of deciding if a quadratic equation over a free group is satisfiable was shown to be decidable. In addition it was shown in [17], [9], and [10] that if n , the number of variables, is fixed, then deciding if a standard quadratic equation has a solution can be done in time which is polynomial in the sum of the lengths of the coefficients. These results imply that the problem is solvable in at most exponential time. In [11] it was shown that the problem of deciding if a quadratic equation over a free group is satisfiable is NP-complete. We will improve on this by proving that in a free group, the length of the value of each variable in a minimal solution of a standard quadratic equation is bounded by $2s$ for orientable equation and by $3s^2$ for non-orientable equation, where s is the sum of the lengths of the coefficients.

Definition 0. Let G be a group and w be an element of its commutator subgroup. We define the *orientable genus* $g(w)$ of w as the least positive integer g such that w is a product of g commutators in G . We define the *non-orientable genus* $g(w)$ of w as the least positive integer n such that w is a product of g squares in G .

Theorem 1. Let C be an orientable (resp., non-orientable) word of genus g in a free group F . Then C can be presented in the form $[a_1, b_1] \dots [a_g, b_g]$ (resp. $d_1^2 \dots d_k^2 [a_1, b_1] \dots [a_m, b_m]$ with $k + 2m = g$), where $|a_i| < 2|C|$, $|b_i| < 2|C|$, $|d_i| < 2|C|$ for $i = 1, \dots, g$.

If C is non-orientable then C can be represented as a product of squares $a_1^2 \dots a_g^2$ with $|a_i| \leq 3|C|^2$.

An *orientable quadratic set of words* is a set of cyclic words w_1, w_2, \dots, w_k (a cyclic word is the orbit of a linear word under cyclic permutations) in some alphabet $a_1^{\pm 1}, a_2^{\pm 1}, \dots$ of letters a_1, a_2, \dots and their inverses $a_1^{-1}, a_2^{-1}, \dots$ such that

- (i) if a_i^ϵ appears in w_i (for $\epsilon \in \{\pm 1\}$) then $a_i^{-\epsilon}$ appears exactly once in w_j ,
- (ii) the word w_i contains no cyclic factor (subword of cyclically consecutive letters in w) of the form $a_l a_l^{-1}$ or $a_l^{-1} a_l$ (no cancellation),

To define a *quadratic set of words* we replace condition (i) by the condition that each $a_i^{\pm 1}$ appears in the set of words w_1, w_2, \dots, w_k exactly twice. Quadratic sets of words were first defined in [14], p.60.

The *genus* of a quadratic set of words is defined as the sum of genera of the surfaces obtained from k discs with words w_1, w_2, \dots, w_k on their boundaries when we identify the edges labeled by the same letters.

The Proposition below follows from Olshanskii's theorem (Theorem 4), described later:

Proposition The following two conditions on a set $\{C_1, \dots, C_k\}$ of elements of a free group F are equivalent:

- (a) The system $\{W_i = C_i, i = 1, \dots, k\}$ has a solution in F where $\{W_i\}$ (orientable or non-orientable) quadratic set of words of genus g ;
- and

- (b) The standard (resp. orientable or non-orientable) quadratic equation of genus g with coefficients $\{C_1, \dots, C_k\}$ has a solution in F .

Proof By Olshanskii's theorem, (b) implies that there is a collection of discs D_1, \dots, D_m with boundaries labelled by a quadratic set of words W_1, \dots, W_m in some variables $P = \{p_1, \dots, p_n\}$, and there is a mapping $\bar{\psi}: P \rightarrow (A \cup A^{-1})^*$ such that upon substitution, the coefficients C_1, \dots, C_{m-1} and C can be read without cancellations around the boundaries of D_1, \dots, D_{m-1} and D_m , respectively. Then $\bar{\psi}(p_1), \dots, \bar{\psi}(p_n)$ is a solution of the system in $W_1 = C_1, \dots, W_m = C$. Let g_0, \dots, g_l be genera of the surfaces obtained from m discs with words W_1, W_2, \dots, W_k on their boundaries when we identify the edges labeled by the same letters. Inequalities in (iii) imply that $\sum_{i=0}^l g_i \leq g$. If this inequality is strict, and $g - \sum_{i=0}^l g_i = r > 0$, we can consider instead of $W_1 = C_1$ equation $W_1[p_{n+1}, p_{n+2}] \dots [p_{n+2r-1}, p_{n+2r}] = C_1$ and define $\bar{\psi}(p_j) = 1$ for $j = n+1, \dots, n+2r$. Then $W_1[p_{n+1}, p_{n+2}] \dots [p_{n+2r-1}, p_{n+2r}], W_2, \dots, W_m$ has genus g . This proves (a).

Now, suppose we have (a), and W_i are equations in variables P . If some letter $p_i \in P$ is contained in different equations we can express it from one equation and substitute in the other. For example, $W_1 = W_{11}p_1W_{12} = C_1$ and $W_2 = W_{21}p_1^{-1}W_{22} = C_2$ become one equation

$$W_{21}W_{12}C_1^{-1}W_{11}W_{22} = C_2$$

which can be rewritten as $W_{21}W_{12}W_{11}W_{22} = C_1^{(W_{21}W_{12})^{-1}}C_2$. If we label the edges of two polygons D_1 and D_2 by W_1 and W_2 , then the left hand side of this equation will be written on the boundary of a polygon obtained by identifying the edges labeled by p_i and removing them. We continue this procedure until there is no letter p_i contained in two different equations. Taking the inverses of both sides of each equation, we obtain a system $L_i(P) = R_i, j = 1, \dots, s$ such that both appearances of every letter from P are contained in one word L_i and each R_i is a product of some conjugates of coefficients C_j^{-1} (if the set of words is orientable) or $C_j^{\pm 1}$ (if the set of words is non-orientable). The sum of topological genera (see definition in Section 3) of words L_i is the sum of genera of the surfaces obtained from s discs with words L_1, \dots, L_s on their boundaries when we identify the edges labeled by the same letters. By construction, the same surfaces are obtained from k discs with words W_1, \dots, W_k on their boundaries. The sum of their genera is g . Topological and algebraic genus of each word L_i is the same. Therefore, the left side of the equation

$$L_1 \dots L_s = R_1 \dots R_s$$

has algebraic genus g . In the orientable case this proves (b). In the non-orientable case we have to change variables and replace negative powers of C_i 's by positive powers. Changing variables we can assume that the left-hand side is a product of g squares. We can also assume that the right-hand side is a product of two parts. The first part is the product of conjugates of negative powers of C_i 's and the second part is the product of conjugates of positive powers of C_i 's. Therefore

$$L_1 \dots L_s = R_1 \dots R_s$$

can be transformed to the form $x_1^2 \dots x_g^2 = P_1 P_2$, where $P_1 = \prod C_{i_j}^{-Z_j}$ and $P_2 = \prod C_{i_l}^{Z_l}$. We will write it as

$$x_1^2 \dots x_g P_2^{-1} P_2 x_g P_2^{-1} = P_1,$$

make a substitution $\bar{x}_g = x_g P_2^{-1}$ and re-write as

$$x_1^2 \dots \bar{x}_g^2 P_2^{\bar{x}_g} P_1^{-1} = 1.$$

Now conjugates of all C_i 's appear in positive exponents, conjugating again we can put them in the right order. Since the system in (a) has a solution, the standard quadratic equation also has a solution, and (b) is proved.

Theorem 2. Let W_1, \dots, W_k be an orientable quadratic set of words of genus g , and C_1, \dots, C_k be elements of a free group F such that the system $W_i = C_i, i = 1, \dots, k$ has a solution in F and $\sum_{i=1}^k |C_i| = s$ in F . Then some product of conjugates of C_i 's in any order can be presented as a product of at most g commutators of elements in F with lengths strictly less than $2s$, and conjugating elements also have length bounded by $2s$.

Let W_1, \dots, W_k be a non-orientable quadratic set of words of genus g , and C_1, \dots, C_k be elements of a free group F such that the system $W_i = C_i, i =$

$1, \dots, k$ has a solution in F and $\sum_{i=1}^k |C_i| = s$ in F . Then some product of conjugates of C_i 's in any order can be presented as a product of at most g squares $a_1^2 \dots a_g^2$ with $|a_i| \leq 12s^4$ and conjugating elements have length bounded by $2s^2$.

Theorem 3. Let h be an orientable word of genus g in a hyperbolic group Γ . Let M be the number of elements in Γ represented by words of length at most 4δ in $F(X)$ (δ is the hyperbolicity constant), $l = \delta(\log_2(12g-6)+1)$. Then h can be presented in a form $[a_1, b_1] \dots [a_g, b_g]$, where $|a_i| < 2|h| + 3(12g-6)(12l+M+4)$, $|b_i| < 2|h| + 3(12g-6)(12l+M+4)$ for $i = 1, \dots, g$.

2 Quadratic equations

A quadratic equation E with variables $\{x_i, y_i, z_j\}$ and non-trivial coefficients $\{C_i, C\} \in F(A)$ is said to be in *standard form* if its coefficients are expressed as freely and cyclically reduced words in A^* and E has either the form:

$$\left(\prod_{i=1}^g [x_i, y_i] \right) \left(\prod_{j=1}^{m-1} z_j^{-1} C_j z_j \right) C = 1 \text{ or } \left(\prod_{i=1}^g [x_i, y_i] \right) C = 1 \quad (1)$$

where $[x, y] = x^{-1}y^{-1}xy$, in which case we say it is *orientable* or it has the form

$$\left(\prod_{i=1}^g x_i^2 \right) \left(\prod_{j=1}^{m-1} z_j^{-1} C_j z_j \right) C = 1 \text{ or } \left(\prod_{i=1}^g x_i^2 \right) C = 1 \quad (2)$$

in which case we say it is non-orientable. The *genus* of a quadratic equation is the number g in (1) and (2) and m is the number of coefficients. If $g = 0$ then we will define E to be orientable. If E is a quadratic equation we define its *reduced Euler characteristic*, $\overline{\chi}$ as follows:

$$\overline{\chi}(E) = \begin{cases} 2 - 2g & \text{if } E \text{ is orientable} \\ 2 - g & \text{if } E \text{ is not orientable} \end{cases}$$

We finally define the *length* of a quadratic equation E to be

$$\text{length}(E) = |C_1| + \dots + |C_{m-1}| + C + 2(\text{number of variables})$$

It is a well known fact that an arbitrary quadratic equation over a free group can be brought to a standard form in time polynomial in its length.

Corollary 1 of Theorem 2. Let $E(X) = 1$ be a standard consistent quadratic equation in a free group F with the set of variables X . For the orientable equation there exists a solution $\phi : F(X) * F \rightarrow F$ such that for each $x \in X$, $|\phi(x)| \leq 2s$, where s is the sum of the lengths of the coefficients. For the non-orientable equation, $|\phi(x)| \leq 12s^4$,

Let G be a group. Let $G[X] = G * F(X)$, where $X = \{x_1, \dots, x_n\}$. Let $S(X) = 1$ be a system of equations over G , that is, $S \subset G[X]$. By $V_G(S)$ denote

the set of all solutions in G of the system $S(X) = 1$, it is called the algebraic set defined by S . $V_G(S)$ uniquely corresponds to the normal subgroup

$$R(S) = \{T(x) \in G[X] \mid \forall g \in G^n (S(g) = 1 \rightarrow T(g) = 1)\}$$

of the group $G[X]$. The quotient group

$$G_{R(S)} = G[X]/R(S)$$

is called the coordinate group of the system $S(X) = 1$.

Let G be a group with a generating set A . A system of equations $S = 1$ is called *triangular quasi-quadratic* (shortly, TQ) over G if it can be partitioned into the following subsystems

$$\begin{aligned} S_1(X_1, X_2, \dots, X_n, A) &= 1, \\ S_2(X_2, \dots, X_n, A) &= 1, \\ &\dots \\ S_n(X_n, A) &= 1 \end{aligned}$$

where for each i one of the following holds:

- 1) S_i is quadratic in variables X_i ;
- 2) If we denote $G_i = G_{R(S_i, \dots, S_n)}$ for $i = 1, \dots, n$, and put $G_{n+1} = G$. $S_i = \{[y, z] = 1, [y, v] = 1, v \in C(u) \mid y, z \in X_i\}$ where u is a group word in $X_{i+1} \cup \dots \cup X_n \cup A$, and $C(u)$ the centraliser of u in G_{i+1} . In this case we say that $S_i = 1$ corresponds to an extension of a centraliser;
- 3) $S_i = \{[y, z] = 1 \mid y, z \in X_i\}$;
- 4) S_i is the empty equation.

The TQ system $S = 1$ is called *non-degenerate* (shortly, NTQ) if the following condition hold:

- 5) each system $S_i = 1$, where X_{i+1}, \dots, X_n are viewed as the corresponding constants from G_{i+1} (under the canonical maps $X_j \rightarrow G_{i+1}$, $j = i + 1, \dots, n$) has a solution in G_{i+1} ;

Corollary 2 of Theorem 2. Let $E(X) = 1$ be an NTQ system in the standard form (all quadratic equations are in the standard form) in a free group F with the set of variables X and with l levels. Let s_i be the total length of the coefficients on the level i . There exists a solution $\phi : F(X) * F \rightarrow F$ such that for each $x \in X$, $|\phi(x)| \leq 12^{2l} s_1^4 \dots s_l^4$.

3 Orientable and non-orientable Wicks forms

Definition 1. An *orientable Wicks form* is a cyclic word w (a cyclic word is the orbit of a linear word under cyclic permutations) in some alphabet $a_1^{\pm 1}, a_2^{\pm 1}, \dots$ of letters a_1, a_2, \dots and their inverses $a_1^{-1}, a_2^{-1}, \dots$ such that

- (i) if a_i^ϵ appears in w (for $\epsilon \in \{\pm 1\}$) then $a_i^{-\epsilon}$ appears exactly once in w ,
- (ii) the word w contains no cyclic factor (subword of cyclically consecutive letters in w) of the form $a_i a_i^{-1}$ or $a_i^{-1} a_i$ (no cancellation),
- (iii) if $a_i^\epsilon a_j^\delta$ is a cyclic factor of w then $a_j^{-\delta} a_i^{-\epsilon}$ is not a cyclic factor of w (substitutions of the form $a_i^\epsilon a_j^\delta \mapsto x$, $a_j^{-\delta} a_i^{-\epsilon} \mapsto x^{-1}$ are impossible).

An orientable Wicks form w is an element of the commutator subgroup when considered as an element in the free group G generated by a_1, a_2, \dots . We define the *algebraic genus* $g_a(w)$ of w as the least positive integer g_a such that w is a product of g_a commutators in G .

The *topological genus* $g_t(w)$ of an orientable Wicks form w is defined as the topological genus of the orientable compact connected surface obtained by labelling and orienting the edges of a $2e$ -gon (which we consider as a subset of the oriented plane) according to w and by identifying the edges in the obvious way.

Definition 2. A *non-orientable Wicks form* is a cyclic word w (a cyclic word is the orbit of a linear word under cyclic permutations) in some alphabet $a_1^{\pm 1}, a_2^{\pm 1}, \dots$ of letters a_1, a_2, \dots and their inverses $a_1^{-1}, a_2^{-1}, \dots$ such that

- (i) if a_i^ϵ appears in w (for $\epsilon \in \{\pm 1\}$) then $a_i^{\pm \epsilon}$ appears exactly once in w , and there is at least one letter which appears with the same exponent,
- (ii) the word w contains no cyclic factor (subword of cyclically consecutive letters in w) of the form $a_i a_i^{-1}$ or $a_i^{-1} a_i$ (no cancellation),
- (iii) if $a_i^\epsilon a_j^\delta$ is a cyclic factor of w then $a_j^{-\delta} a_i^{-\epsilon}$ or another appearance of $a_i^\epsilon a_j^\delta$ is not a cyclic factor of w .

A non-orientable Wicks form w is an element of the subgroup of squares when considered as an element in the free group G generated by a_1, a_2, \dots . We define the non-orientable *algebraic genus* $g_a(w)$ of w as the least positive integer g_a such that w is a product of g_a squares in G .

The *topological genus* $g_t(w)$ of a non-orientable Wicks form w is defined as the topological genus of the non-orientable compact connected surface obtained by labelling and orienting the edges of a $2e$ -gon according to w and by identifying the edges in the obvious way.

Remark The algebraic and the topological genus of an orientable Wicks form coincide (cf. [3, 5]). The same is true for non-orientable Wicks forms.

We define the *genus* $g(w)$ of a Wicks form w by $g(w) = g_a(w) = g_t(w)$.

Consider the orientable compact surface S associated to an orientable (non-orientable) Wicks form w . This surface carries an immersed graph $\Gamma \subset S$ such

that $S \setminus \Gamma$ is an open polygon with $2e$ sides (and hence connected and simply connected). Moreover, conditions (ii) and (iii) on Wicks form imply that Γ contains no vertices of degree 1 or 2 (or equivalently that the dual graph of $\Gamma \subset S$ contains no faces which are 1-gons or 2-gons). This construction works also in the opposite direction: Given a graph $\Gamma \subset S$ with e edges on an orientable (non-orientable) compact connected surface S of genus g such that $S \setminus \Gamma$ is connected and simply connected, we get an orientable(non-orientable) Wicks form of genus g and length $2e$ by labelling and orienting the edges of Γ and by cutting S open along the graph Γ . The associated orientable(non-orientable) Wicks form is defined as the word which appears in this way on the boundary of the resulting polygon with $2e$ sides. We identify henceforth orientable(non-orientable) Wicks forms with the associated immersed graphs $\Gamma \subset S$, speaking of vertices and edges of orientable (non-orientable) Wicks form.

The formula for the Euler characteristic

$$\chi(S) = 2 - 2g = v - e + 1$$

in orientable and

$$\chi(S) = 2 - g = v - e + 1$$

in non-orientable case (where v denotes the number of vertices and e the number of edges in $\Gamma \subset S$) shows that an orientable Wicks(non-orientable) form of genus g has at least length $4g(2g)$ (the associated graph has then a unique vertex of degree $4g$ and $2g$ edges) and at most length $6(2g-1)(6(g-1))$ (the associated graph has then $2(2g-1)(2(g-1))$ vertices of degree three and $3(2g-1)((3(g-1))$ edges).

We call an orientable Wicks form of genus g *maximal* if it has length $6(2g-1)$ in orientable and $6(g-1)$ in non-orientable case.

A vertex V (with oriented edges a, b, c pointing toward V) is *positive* if

$$w = ab^{-1} \dots bc^{-1} \dots ca^{-1} \dots \quad \text{or} \quad w = ac^{-1} \dots cb^{-1} \dots ba^{-1} \dots$$

and V is *negative* if

$$w = ab^{-1} \dots ca^{-1} \dots bc^{-1} \dots \quad \text{or} \quad w = ac^{-1} \dots ba^{-1} \dots ab^{-1} \dots$$

Let V be a negative vertex of an orientable maximal Wicks form of genus $g > 1$. There are three possibilities, denoted configurations of type α , β and γ (see Figure 1) for the local configuration at V .

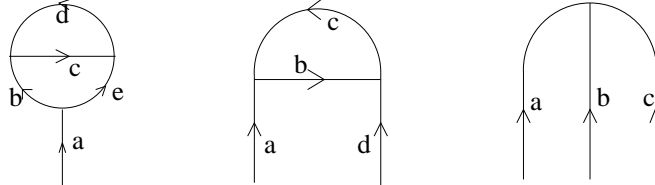


Figure 1.

Type α . The vertex V has only two neighbours which are adjacent to each other. This implies that w is of the form

$$w = w_1 x_1 a b c d b^{-1} e c^{-1} d^{-1} e^{-1} a^{-1} x_2 w_2 x_2^{-1} x_1^{-1} w_3$$

(where w_1, w_2, w_3 are subfactors of w) and w is obtained from the maximal orientable Wicks form

$$w' = w_1 x w_2 x^{-1} w_3$$

of genus $g-1$ by the substitution $x \mapsto x_1 a b c d b^{-1} e c^{-1} d^{-1} e^{-1} a^{-1} x_2$ and $x^{-1} \mapsto x_2^{-1} x_1^{-1}$ (this construction is called the α -construction in [19]).

Type β . The vertex V has two non-adjacent neighbours. The word w is then of the form

$$w = w_1 x_1 a b c a^{-1} x_2 w_2 y_1 d b^{-1} c^{-1} d^{-1} y_2 w_3$$

(where perhaps $x_2 = y_1$ or $x_1 = y_2$, see [19] for all the details). The word w is then obtained by a β -construction from the word $w' = w_1 x w_2 y w_3$ which is an orientable maximal Wicks form of genus $g-1$.

Type γ . The vertex V has three distinct neighbours. We have then

$$w = w_1 x_1 a b^{-1} y_2 w_2 z_1 c a^{-1} x_2 w_3 y_1 b c^{-1} z_2 w_4$$

(some identifications among x_i, y_j and z_k may occur, see [19] for all the details) and the word w is obtained by a so-called γ -construction from the word $w' = w_1 x w_3 y w_2 z w_4$, where $x = x_1 x_2, y = y_1 y_2, z = z_1 z_2$.

For each of α, β or γ an inverse transformation is well defined (see [19] for the details), so we can associate a Wicks form of genus $g-1$ to a Wicks form of genus g and a given negative vertex.

Definition 3. We call the application which associates to an orientable maximal Wicks form w of genus g with a chosen negative vertex V the orientable maximal Wicks form w' of genus $g-1$ defined as above the *reduction* of w with respect to the negative vertex V .

An inspection of figure 1 shows that reductions with respect to vertices of type α or β are always paired since two doubly adjacent vertices are negative, of the same type (α or β) and yield the same reductions.

The above constructions of type α, β and γ can be used for a recursive construction of all orientable maximal Wicks forms of genus $g > 1$ ([19]).

Let a be an edge of a non-orientable maximal genus g Wicks form w appearing with the same exponent. Without loss of generality we can assume that $w = v_1 a v_2 a v_3$. The word w' obtained from $v_1 v_2^{-1} v_3$ by simplifications and reductions is either non-orientable genus $g-1$ Wicks form or orientable genus m Wicks form, where $g = 2m + 1$ (see [20] for details).

Definition 4. We will say that the word w' is obtained from a non-orientable Wicks form w by reduction with respect to the edge a .

4 Proof of the Theorem 1

We treat the orientable case first. Since the word C has genus g , it can be obtained from a Wicks form U of genus g by a non-cancelling substitution (it means, that all letters of U are replaced by words from a free group such that no cancellations occur). Any Wicks form of genus g can be obtained from the one of maximal length by substituting some letters by empty words ([19]), that is why without loss of generality we can assume, that U has maximal length $12g - 6$ (some letters in the Wicks form can be thought as replaced by empty words). By [19], any Wicks form of maximal length always contains a negative vertex. If V is a vertex with leaving edges a, b, c , then

$$C = u_1 a^{-1} b u_2 c^{-1} a u_3 b^{-1} c u_4,$$

where the vertex V appear as a part of one of the subgraphs on the figure 1 and some of subwords u_1, u_2, u_3, u_4 may be empty according to this. Now we will rewrite C using properties of U .

$$\begin{aligned} C &= u_1 a^{-1} b u_2 c^{-1} a u_3 b^{-1} c u_4 = \\ &u_1 a^{-1} b u_2 c^{-1} a u_3 b^{-1} c u_2^{-1} u_2 u_4 = \\ &u_1 u_3 u_3^{-1} a^{-1} b u_2 c^{-1} a u_3 b^{-1} c u_2^{-1} u_2 u_4 = \\ &u_1 u_3 u_3^{-1} a^{-1} b u_2 c^{-1} a u_3 b^{-1} c u_2^{-1} u_3^{-1} u_1^{-1} u_1 u_3 u_2 u_4 = \\ &u_1 u_3 [u_3^{-1} a^{-1} c u_2^{-1}, u_2 c^{-1} b] u_3^{-1} u_1^{-1} u_1 u_3 u_2 u_4 = \\ &[u_1 u_3 u_3^{-1} a^{-1} c u_2^{-1} u_3^{-1} u_1^{-1}, u_1 u_3 u_2 c^{-1} b u_3^{-1} u_1^{-1}] u_1 u_3 u_2 u_4 = \\ &[u_1 a^{-1} c u_2^{-1} u_3^{-1} u_1^{-1}, u_1 u_3 u_2 c^{-1} b u_3^{-1} u_1^{-1}] u_1 u_3 u_2 u_4. \end{aligned}$$

Denote $u_1 a^{-1} c u_2^{-1} u_3^{-1} u_1^{-1}$ by a_1 and $u_1 u_3 u_2 c^{-1} b u_3^{-1} u_1^{-1}$ by b_1 . It is easy to see, that $|a_1| \leq 2|C|$ and $|b_1| \leq 2|C|$. Now we can see that the word $u_1 u_3 u_2 u_4$ has genus $g - 1$ since it can be obtained by a non-cancelling substitution from a Wicks form U , which is obtained from U by reduction with respect to the vertex V .

The word $C = u_1 a^{-1} b u_2 c^{-1} a u_3 b^{-1} c u_4$ is obtained from $u_1 u_3 u_2 u_4$ by transformation γ , if all $|u_1| + |u_4|$, $|u_2|$, $|u_3|$ are not equal 0. Indeed, the word $u_1 u_3 u_2 u_4$ is corresponding to $w' = w_1 x w_3 y w_2 z w_4$ in the description of the transformation γ as follows: u_1 is corresponding to $w_1 x_1$, u_3 is corresponding to $x_2 w_3 y_1$, u_2 is corresponding to $y_2 w_2 z_1$, u_4 is corresponding to $z_2 w_4$, a is corresponding a^{-1} , b is corresponding to b^{-1} , c is corresponding to c^{-1} , and $x_1 x_2$ is replaced by $x, y_1 y_2$ is replaced by y , and $z_1 z_2$ is replaced by y .

C is obtained by transformation β if exactly one of $|u_1| + |u_4|, |u_2|, |u_3|$ is equal to 0, otherwise C is obtained from $u_1 u_3 u_2 u_4$ by transformation α .

In the case of β we can assume without loss of generality that $|u_2| = 0$, then $C = u_1 a b c a^{-1} u_3 b^{-1} c^{-1} u_4$. In this case $u_3 = u'_2 d, u_4 = d^{-1} u'_3$, where d can be empty, and $C = [u_1 a b d^{-1} (u'_2)^{-1}, u_1 u'_2 d c a^{-1} u_1^{-1}] u_1 u'_2 u'_3$. The word $u_1 u'_2 u'_3$ corresponds to $w' = w_1 x w_2 y w_3$ as follows: u_1 corresponds to $w_1 x_1$, u'_2

corresponds to $x_2w_2y_1$, u'_3 corresponds to y_2w_3 , and x_1x_2 is replaced by x and y_1y_2 is replaced by y .

The easiest case is the transformation *alpha* and without loss of generality we will assume that $C = u_1abcd b^{-1}ec^{-1}d^{-1}e^{-1}a^{-1}u_4$ and the negative vertex is the vertex that the edges b and d are incoming edges, and c is the leaving edge, $C = [u_1abce^{-1}a^{-1}u_1^{-1}, u_1aedb^{-1}a^{-1}u_1^{-1}]u_1u_2$. The word u_1u_2 corresponds to $w' = w_1xw_2x^{-1}w_3$ by $u_1 = w_1x_1$, $u_2 = x_2w_2x^{-1}w_3$ where x_1x_2 is replaced by x .

For details on transformations α , β , γ see [19]

Continuing by induction, we get the statement of the Theorem 1 in the orientable case.

Now let C be a non-orientable word of genus g . Then it can be obtained from a non-orientable Wicks form w by a non-cancelling substitution. We rewrite C using the fact that w has a letter appearing with the same exponent:

$$C = v_1av_2av_3 = (v_1av_2v_1^{-1})^2v_1v_2^{-1}v_3.$$

Replacing $v_1av_2v_1^{-1}$ with x_1 it is easy to see, that $|x_1| \leq 2|C|$ and $|v_1v_2^{-1}v_3| \leq |C|$. If $v_1v_2^{-1}v_3$ is orientable, then we use the orientable case of the Theorem 1, which we've just proved. If $v_1v_2^{-1}v_3$ is non-orientable, then it can be obtained from non-orientable Wicks form w' and we proceed as above.

Let C be a non-orientable word of genus g . It is easy to see, that a product of a square and a commutator can be rewritten as a product of three squares in the following way: $x^2[a, b] = (x^2abx^{-1})^2(xb^{-1}a^{-1}x^{-1}a^{-1}x^{-1})^2(xa)^2$. We have just proved that $C = d_1^2 \dots d_k^2[a_1, b_1] \dots [a_m, b_m]$, where $g = k + 2m$ and $|a_i| < 2|C|, |b_i| < 2|C|, |d_i| < 2|C|$ for $i = 1, \dots, g$. Using our rewriting equality, we get that $|a_i| < 6m2|C| \leq 6g|C| \leq 3|C|^2$.

Theorem 1 is proved.

5 Ol'shanskii's result

The following is proved in [17].

Theorem 4 Let E be a quadratic equation over $F(A)$ in standard form. If $g = 0, m = 2$, or E is not orientable and $g = 1, m = 1$ then we set $N = 1$. Otherwise we set $N \leq 3(m - \bar{\chi}(E))$. E has a solution if and only if for some $n \leq N$;

- (i) there is a set $P = \{p_1, \dots, p_n\}$ of variables and a collection of m discs D_1, \dots, D_m such that,
- (ii) the boundaries of these discs are circular 1-complexes with directed and labelled edges such that each edge has a label in P and each $p_j \in P$ occurs exactly twice in the union of boundaries;
- (iii) Let $\bar{\chi}(E) = 2 - 2g$ for orientable surface and $\bar{\chi}(E) = 2 - g$ for non orientable surface. If we glue the discs together by edges with the same label, respecting the edge orientations, then we will have a collection $\Sigma_0, \dots, \Sigma_l$

of closed surfaces and the following inequalities: if E is orientable then each Σ_i is orientable and

$$\left(\sum_{i=0}^l \chi(\Sigma_i) \right) - 2l \geq \bar{\chi}(E)$$

if E is non-orientable either at least one Σ_i is non-orientable and

$$\left(\sum_{i=0}^l \chi(\Sigma_i) \right) - 2l \geq \bar{\chi}(E)$$

or, each Σ_i is orientable and

$$\left(\sum_{i=0}^l \chi(\Sigma_i) \right) - 2l \geq \bar{\chi}(E) + 2$$

and

- (iv) there is a mapping $\bar{\psi} : P \rightarrow (A \cup A^{-1})^*$ such that upon substitution, the coefficients C_1, \dots, C_{m-1} and C can be read without cancellations around the boundaries of D_1, \dots, D_{m-1} and D_m , respectively; and finally that
- (v) if E is orientable the discs D_1, \dots, D_m can be oriented so that w_i is read clockwise around ∂D_i and d is read clockwise around ∂D_m , moreover all these orientations must be compatible with the gluings.

Proof. It is shown in Sections 2.4 [17] that the solvability of a quadratic equation over $F(A)$ coincides with the existence of a *diagram* Δ over $F(A)$ on the appropriate surface Σ with boundary. This diagram may not be *simple*, so via surgeries we produce from Σ a finite collection of surfaces $\Sigma_1, \dots, \Sigma_l$ with induced simple diagrams $\Delta_1, \dots, \Delta_l$ which we can recombine to get back Σ and Δ . So existence of a diagram Δ on Σ is equivalent to existence of a collection of simple diagrams Δ_i on surfaces Σ_i such that the inequalities involving Euler characteristics given in the statement of the Theorem are satisfied.

In Section 2.3 of [17] the bounds on n are proved. It is also shown in that section that if one can glue discs together as described in the statement of the Theorem with the condition on the boundaries, then there exist simple diagrams Δ_i on surfaces Σ_i . \square

6 Quadratic sets of words

An *orientable quadratic set of words* is a set of cyclic words w_1, w_2, \dots, w_k (a cyclic word is the orbit of a linear word under cyclic permutations) in some alphabet $a_1^{\pm 1}, a_2^{\pm 1}, \dots$ of letters a_1, a_2, \dots and their inverses $a_1^{-1}, a_2^{-1}, \dots$ such that

- (i) if a_i^ϵ appears in w_i (for $\epsilon \in \{\pm 1\}$) then $a_i^{-\epsilon}$ appears exactly once in w_j ,

- (ii) the word w_i contains no cyclic factor (subword of cyclically consecutive letters in w) of the form $a_l a_l^{-1}$ or $a_l^{-1} a_l$ (no cancellation),

The *genus* of a quadratic set of words is defined as the sum of genera of the surfaces obtained from discs k with words w_1, w_2, \dots, w_k on their boundaries.

Proof of Theorem 2. Let's consider an orientable case first.

Let W_1, \dots, W_k be an orientable quadratic set of words in an alphabet $\mathcal{A} = \{a_1^{\pm 1}, a_2^{\pm 1}, \dots\}$ of genus g , and C_1, \dots, C_k be elements of a free group F such that the system $W_i = C_i$, $i = 1, \dots, k$ has a non-cancelable solution $\phi : F(\mathcal{A}) \rightarrow F$ in F and $\sum_{i=1}^k |C_i| = s$ in F .

We will transform the system $W_i = C_i$, $i = 1, \dots, k$ to one equality.

Suppose there is a letter a_1 contained in two equations $C_1 = W_1 = U_1 a_1 U_2$ and $C_2 = W_2 = U_3 a_1^{-1} U_4$. This corresponds to the substitution of a_1 from the first equation to the second. We obtain $1 = C_2 U_4^{-1} U_1^{-1} C_1 U_2^{-1} U_3^{-1}$. Until there is a letter, contained in two different equations, we rewrite these two equations as one. As soon as both appearances of every letter from \mathcal{A} are contained in one word $L_j = 1$, we write all L_j next to each other in any order and obtain a word W quadratic in a_1, \dots, a_n containing each C_i only once. We claim, that some conjugates of C_i in any order can be presented as a product of at most g commutators of elements with lengths strictly less than $2s$, and conjugating elements also have length bounded by $2s$.

Without loss of generality we can assume that the last coefficient C_k of the standard form appears at the right end of the word W . Now let $W = M_3 C_{k-1} M_4 C_k$. Then by conjugating C_{k-1} by M_4^{-1} we get $W = M_3 C_{k-1}^{M_4^{-1}} C_k$.

The length of $\phi(W)$ is bounded by $2s$ because the sum of length of all $\phi(a_i)$'s taken twice is not larger than s , and the sum of C_i' 's is bounded by s as well. This proves that conjugating element M_1 no longer than $2s$. Continuing by induction, we get that the product of some conjugates of C_i' 's equals to a quadratic word of genus g . (We have genus g since the initial quadratic set of words had genus g , this has been proved in the proof of Proposition, (a) \rightarrow (b)). This word can be presented as a Wicks form of genus g . Now by using the statement of the Theorem 1, we represent the product of conjugates as the product of commutators of elements of length not larger than $2s$.

In the non-orientable case there are two cases: when the letter from \mathcal{A} which we use to bring two equations together, appears with different exponents or with the same. If a letter appears in two equations with different exponents, we treat these equations the same as in the orientable case. But if a letter saying a_t appears in W_h and W_p with same exponent, and W_h and W_p don't contain a letter with different exponents, namely $W_h = X_1 a_t X_2 = C_h$ $W_p = X_3 a_t X_4 = C_p$, then the equation obtained by combining $W_h = X_1 a_t X_2 = C_h$ and $W_p = X_3 a_t X_4 = C_p$ contains C_p and C_h with different exponents. So, our final word W , obtained by bringing all the equations of the system together, can contain the coefficients with exponent -1 , such that every of these coefficients C_h 's appears as C_h^{-1} in W . Let a_t be a letter with appears with the same exponent in W . By conjugation, we can bring all C_h 's next to a_t such that the length of $\phi(W')$ is not more than $2s$. $W' = Y_1 a_t C_{h_1}^{-1} C_{h_2}^{-1} \dots C_{h_r}^{-1} Y_2 a_t Y_3$, where

$\tilde{C}_{h_1}^{-1}, \tilde{C}_{h_2}^{-1}, \dots, \tilde{C}_{h_r}^{-1}$ are conjugates of $C_{h_1}^{-1}, C_{h_2}^{-1}, \dots, C_{h_r}^{-1}$. Without loss of generality we can assume, that $r \leq k/2$, where k is the number of equations in the system (otherwise, we take the inverse equation). Next to the second appearance of a_t we insert $\tilde{C}_{h_1}^{-1} \tilde{C}_{h_2}^{-1} \dots \tilde{C}_{h_r}^{-1} (\tilde{C}_{h_1}^{-1} \tilde{C}_{h_2}^{-1} \dots \tilde{C}_{h_r}^{-1})^{-1}$. Substituting $a_t \tilde{C}_{h_1}^{-1} \tilde{C}_{h_2}^{-1} \dots \tilde{C}_{h_r}^{-1}$ by a'_t , we get a new equality, which image under ϕ has length not more then $4rs \leq 2ks$ (since $r \leq k/2$) and such that all coefficients are with the same exponent. Proceeding as in the orientable case, we get our equality to the standard form, taking into account the non-orientable case of the Theorem 1. First we obtain the product of the conjugates of coefficients in any order followed by $d_1^2 \dots d_k^2 [a_1, b_1] \dots [a_m, b_m]$ with $k + 2m = g$ and $\phi(a_i), \phi(b_i), \phi(d_i)$ no longer than $4ks$, and then use the same transformation as in the proof of Theorem 1. Since $k \leq s$, we get the statement of Theorem 2.

7 Hyperbolic groups

In this section we will prove Theorem 3.

Let h be a word of genus g in $\Gamma = \langle X | R \rangle$ such that $h = [a_1, b_1] \dots [a_g, b_g]$. The orientable word $U = [A_1, B_1] \dots [A_g, B_g]$ is a genus g orientable Wicks form. Let $L = \{A_1, B_1, \dots, A_g, B_g\}$ and let $\phi : F(L) \rightarrow F(X)$ (where $F(L)$ ($F(X)$) is a free group with basis L (X)) be a homomorphism $\phi(A_i) = a_i$, $\phi(B_i) = b_i$ with $i = 1, \dots, g$. We call this a labelling function for U . Let \mathcal{F} be the set of pairs (U, ϕ) where U is a genus g Wicks form and ϕ is a labelling function for U such that $\phi(U)$ is conjugated to h in Γ .

Consider a pair (W, θ) in which $|\theta(W)|$ is minimal amongst all pairs in \mathcal{F} . Clearly $\theta(E)$ is minimal for each letter E in W . For convenience we shall take \hat{W} to be a cyclic permutation of W such that the last letter of \hat{W} is labelled by a word of length more than $12l + M + 4$. Consider all letters of \hat{W} which have labels greater than $12l + M + 4$ in $F(X)$, all the labels of other letters are shorter than $12l + M + 4$. Corresponding edges in the Cayley graph G of Γ will be called long and short edges respectively.

Consider $\theta(\hat{W})$ as a path in the Cayley graph G . Let C be a word in $F(X)$ which represents geodesic for $\theta(\hat{W})$ and let R be a minimal word such that $h =_\Gamma RCR^{-1}$. We can assume that $2|R| \leq \{\text{length of all short edges}\}$ (otherwise we can take \hat{W} to be a cyclic permutation of W such that the first letter of \hat{W} is labelled by a word of length more than $12l + M + 4$). By Lemma 4.3 of [8] the terminal vertex of each long edge in G is within $5l + M + 3$ of some vertex of C , see Figure 2. (C is corresponding to F in the notation of [8]). Let B be a long edge in \hat{W} which is not the first long edge in the sequence of letters. Since \hat{W} is quadratic, B appears twice, once with exponent 1 and once with exponent -1. First we shall consider the appearance of B with exponent 1. In the sequence of letters of \hat{W} , let $A^{\pm 1}$ be the long edge before B in the sequence such that no long edge appears between $A^{\pm 1}$ and B (note that $A^{\pm 1}$ could be B^{-1}).

Let $\iota(p), \tau(p)$ denote the beginning and the end of the path p .

Lemma 1 [Lemmas 4.3, 4.5 of [8]] There exist vertices u and v on C , such that $d(\tau(\theta(A^{\pm 1})), u), d(\tau(\theta(B)), v) \leq 5l + M + 3$. We can choose u and v such

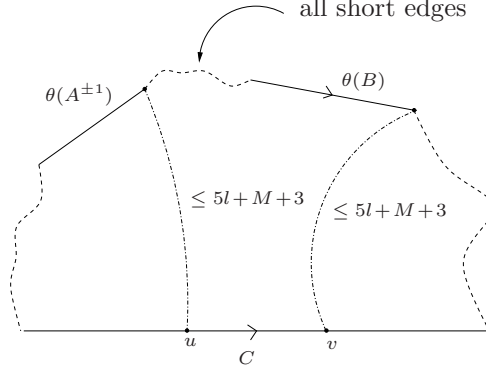


Figure 2.

that $d(\iota(C), u) < d(\iota(C), v)$.

Lemma 2 [Follows from the proof of Lemmas 4.3, 4.5 of [8]] Let $\hat{W} = A_1 A_2 \dots A_n$ and $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ ($i_k = n$), be long edges. We can choose vertices u_1, u_2, \dots, u_k on C such a way that $d(\tau(\theta(A_{i_j})), u_j) \leq 5l + M + 3$ and vertices u_1, u_2, \dots, u_k appear in the path C in the natural order.

Proof. Lemma 4.3 states the existence of a vertex u_j for each long edge A_{i_j} such that $d(\tau(\theta(A_{i_j})), u_j) \leq 5l + M + 3$. The vertices u_j and u_{j+1} chosen in Lemma 4.3 for arbitrary two consecutive long edges, are, actually, chosen such a way that $d(\iota(C), u_j) < d(\iota(C), u_{j+1})$. Indeed, in the proof of Lemma 4.5 it is shown that the assumption $d(\iota(C), u_j) \geq d(\iota(C), u_{j+1})$ implies a contradiction. \square

We now represent accordingly C as $C = D_1 \dots D_k$.

Applying triangle inequalities to long edges, we obtain

$$|\theta A_j| \leq 2(5l + M + 3) + D_j + \{\text{length of short edges between } A_{j-1} \text{ and } A_j\};$$

$$\sum |\theta(A_j)| \leq |C| + 2k(5l + M + 3) + \{\text{length of all short edges}\};$$

Therefore

$$\begin{aligned} |\theta(\hat{W})| &\leq |C| + 2k(5l + M + 3) + 2\{\text{length of all short edges}\} \\ &\leq |C| + (12g - 6)(12l + M + 4). \end{aligned}$$

Indeed, if $2k_1$ is the number of short edges, then $2k + 2k_1 \leq 6(2g - 1)$. Notice now that from the triangle inequality, $|C| \leq |h| + 2|R| \leq |h| + \{\text{length of all short edges}\}$. Therefore,

$$|\theta(\hat{W})| \leq |h| + (18g - 9)(12l + M + 4).$$

And now we complete the proof using Theorem 1.

Acknowledgements We are very thankful to the referee for many useful comments and suggestions.

References

- [1] G.N. Arzhantseva, *On quasiconvex subgroups of word hyperbolic groups*, Geometriae dedicata, 87, 191-208, 2001.
- [2] J.L.Brenner, R.C.Lyndon, *Permutations and cubic graphs*, Pacific Journal of Maths, v.104, 285-315
- [3] M.Culler, *Using surfaces to solve equations in free groups*, Topology, v.20(2), 1981
- [4] J.A.Comerford, L.P.Comerford and C.C.Edmunds, *Powers as products of commutators*, Commun. in Algebra, 19(2), 675-684 (1991)
- [5] L.Comerford, C.Edmunds, *Products of commutators and products of squares in a free group*, Int.J. of Algebra and Comput., v.4(3), 469-480, 1994
- [6] Leo P. Comerford, Jr. and Charles C. Edmunds. Quadratic equations over free groups and free products. *J. Algebra*, 68(2):276-297, 1981.
- [7] V. Diekert, C. Gutierrez, C. Hagenah, The existential theory of equations with rational constraints in free groups is PSPACE-complete, *Information and Computation*, Volume 202, Issue 2 , 1 November 2005, 105-140.
- [8] S.M. Fulthorp, *Squares, Commutators and Genus in Infinite Groups*, PhD thesis, School of Mathematics and Statistics, Newcastle University, 2004; *Genus n forms over Hyperbolic Groups*, ArXiv:10051513v1[MathGR], 2010.
- [9] R. I. Grigorchuk and P. F. Kurchanov. On quadratic equations in free groups. In *Proceedings of the International Conference on Algebra, Part 1 (Novosibirsk, 1989)*, volume 131 of *Contemp. Math.*, pages 159-171, Providence, RI, 1992. Amer. Math. Soc.
- [10] R. I. Grigorchuk and I. G. Lysionok. A description of solutions of quadratic equations in hyperbolic groups. *Internat. J. Algebra Comput.*, 2(3):237-274, 1992.
- [11] O. Kharlampovich, I. Lysenok, A. Myasnikov, N. Touikan, Quadratic equations over free groups are NP-complete, *TOCS (Teor. Comp. Syst.)*, 10, 2008.
- [12] O. Kharlampovich, A. Myasnikov, Implicit function theorem over free groups. *Journal of Algebra*, 290 (2005), 1-203.
- [13] A. I. Malcev. On the equation $zxyx^{-1}y^{-1}z^{-1} = aba^{-1}b^{-1}$ in a free group. *Algebra i Logika Sem.*, 1(5):45-50, 1962.
- [14] R.C.Lyndon, P.E.Schupp, *Combinatorial group theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 89. Springer-Verlag, Berlin-New York, 1977

- [15] J. Macdonald, Compressed words and automorphisms in fully residually free groups, in International Journal of Algebra and Computation, 20, no.3 (2010) 343-355.
- [16] L. Mosher, *A User's Guide to the Mapping Class Group: Once Punctured Surfaces*, DIMACS Series, Vol.25, 1994.
- [17] A. Yu. Olshanskii. Diagrams of homomorphisms of surface groups. *Sibirsk. Mat. Zh.*, 30(6):150–171, 1989.
- [18] A. Yu. Olshanskii, On residualizing homomorphisms and G -subgroups of hyperbolic groups, *IJAC*, 3(4), 1993, 365-409.
- [19] A.A.Vdovina, *Constructing Orientable Wicks Forms and Estimation of Their Number*, *Communications in Algebra* 23 (9), 3205–3222 (1995).
- [20] A. Vdovina, *On the number of nonorientable Wicks forms in a free group*, *Proc. R. Soc. Edinb., Sect. A* 126, No.1, 113-116 (1996).